On the Selection of Arbitrators

Geoffroy de Clippel, Kfir Eliaz, and Brian Knight*

July 6, 2012

Abstract

A key feature of arbitration is the possibility for conflicting parties to partake in the selection of the person who will rule the case. We analyze this problem of the selection of arbitrators from the perspective of implementation theory. Theoretical, empirical and experimental arguments are combined to highlight difficulties with a procedure that is commonly used in practice and to develop and identify better performing procedures.

1 Introduction

Implementation theory studies the design of institutions and procedures for collective decision-making. It aims to find ways of incentivizing participants to select "desirable" outcomes. What is deemed "desirable" varies across situations, and is represented by a social choice rule (SCR) that maps the participants' preferences to subsets of feasible outcomes. When applied to concrete economic environments, this theory helps address a number of important questions. Do prevalent procedures implement the intended SCR? Are there alternative mechanisms? Are there acceptable variants of the SCR that are implementable? How do alternative mechanisms perform when tested with

*Brown University, Department of Economics, Providence, Rhode Island - declippel@brown.edu, kfir_eliaz@brown.edu, brian_knight@brown.edu. We wish to thank Eli Zvuluni of Possible Worlds Inc. for programming the experiment, Melis Kartal and Mark Bernard for running the experiment, CESS at NYU and especially Caroline Madden for invaluable administrative help, Samuel Mencoff, Pantelis Solomon, Ee Cheng Ong and especially Neil Thakral for exceptional research support.

participants facing real stakes? These questions have been studied in a wide variety of contexts including auctions, the provision of public goods, kidney exchange, school choice and choice of medical residency (see the studies surveyed in Kagel (1995), Chen (2008), Roth (2002, 2007) and Kagel and Levin (2011)).

We contribute to this literature by applying implementation theory to a rich class of situations in which individuals must agree on a collective decision, and where monetary transfers are not available. This class includes elections of public officials, committee decisions, selection of committee members, selection of juries for a trial, selection of judges for an appellate court, etc. The present paper focuses on perhaps the simplest problem within this general class: the selection of arbitrators.

Contrary to problems involving committees or a large number of voters, arbitrator selection involves only two parties. Also, contrary to jury selection, which involves the selection of a panel of individuals, the final outcome involves the selection of a single individual, the arbitrator. Selecting an arbitrator is also a case where the assumption of complete information, which underlies many theoretical models, is reasonable. Indeed, most disputes resolved through arbitration occur between parties that have a long-term relationship (e.g., unions and management). In addition, the arbitration agencies provide both parties with the same information about the potential arbitrators. Because arbitrators differ in their fees, their expertise, their past rulings and their delays in reaching a decision, some arbitrators may be ranked above others by both parties (i.e., the parties do not necessarily have completely opposed rankings of all arbitrators).

In addition to being tractable, the problem of selecting an arbitrator is of practical relevance. Arbitration is the most common procedure for resolving disputes without resorting to costly litigation. Having a role in choosing who will rule the case is often cited by participants as one of its main attractive features. Indeed parties dislike facing the risk of being subject to a judge who is not qualified for the case or who is perceived as biased. Hence, the relative appeal of arbitration agencies depends on their ability to assign arbitrators to

cases in a way that best reflects the preferences of both parties.

This paper aims to identify selection mechanisms that satisfy two criteria: (i) a "theoretical" criterion - every equilibrium induced by the mechanism has normatively appealing properties (which we describe shortly), and (ii) an "empirical" criterion - when the mechanism is actually carried out with real incentives, it is likely to generate outcomes that satisfy the desired properties. We compare four mechanisms, two of which are commonly used in practice and two of which are yet to be used. We argue that the latter pair is superior to the former both in theory and in terms of its actual performance in a laboratory setting.

Our analysis proceeds as follows. We first consider the a commonly used procedure for assigning arbitrators, *the Veto-Rank mechanism* (VR).[1] Under this mechanism, two parties receive a list of n (an odd number) potential arbitrators. Each party independently vetoes or removes $\frac{n-1}{2}$ names from the list, and ranks the remaining $\frac{n+1}{2}$ candidates. The selected arbitrator is one with the minimal sum of ranks among candidates who have not been vetoed (ties are resolved via a lottery).

The veto-rank mechanism is appealing *if participants are truthful*, i.e., if they veto their bottom $\frac{n-1}{2}$ candidates and rank the remaining ones truthfully. Specifically, the resulting SCR satisfies two appealing properties: the appointed arbitrator is Pareto efficient and Pareto dominates both parties' median choices (a 'minimal satisfaction' test). However, truth-telling is not always a Nash equilibrium, hence, participants may strictly gain by deviating from truthful behavior.[2] Therefore, actual outcomes may end up violating the above appealing properties. We argue that these concerns apply to *all* simultaneous mechanisms, not just VR. Indeed, Proposition 1 establishes that there is no simultaneous mechanism that Nash implements a SCR that selects Pareto efficient outcomes, which Pareto dominates the parties' median choices. In particular, the SCR derived from truthtelling in VR is *not* Nash

[1]The Supplementary Appendix contains a list of major arbitration agencies that use the veto-rank mechanism to select arbitrators.

[2]If an arbitrator is commonly known among parties to be unqualified for the case, for example, why waste a veto on him if one believes that the other party will veto him?

implementable.

We complement our negative theoretical results with both empirical and experimental evidence that hints at the potential weaknesses of the VR procedure. First, we present empirical evidence from real cases suggesting that in practice parties typically do not have completely opposed preferences. Second, experimental results involving undergraduates confirm that non-truthful behavior occurs in a majority of cases, a significant proportion of which is driven by some strategic motives. If such behavior occurs in a rather abstract environment with small stakes and rather inexperienced subjects, then a fortiori should it happen in real-life conditions with larger stakes and professional players (most often lawyers).

Given the potential problems with simultaneous mechanisms, we turn our attention to three sequential procedures, all of which select an arbitrator that is Pareto efficient and Pareto dominates both parties' median choices.

Alternate Strikes scheme (AS) - In the only sequential procedure that is used in practice, both parties alternatively remove a name from the list of potential arbitrators, and the final remaining option is chosen to be the arbitrator.[3]

Voting by Alternating Offers and Vetoes (VAOV) - In this procedure players take turns in proposing arbitrators.[4] When a proposed arbitrator is rejected by the other party, that arbitrator is removed from the list and the rejecting party then proposes a name from the remaining list. The procedure continues until a proposal is accepted or only one name remains (which is then selected).

Shortlisting (SL) - One party starts the game by selecting $\frac{n+1}{2}$ candidates, and the second party then selects the arbitrator out of that shortlist. This procedure has the appealing feature of having only two rounds, simplifying backwards induction (Binmore et. al., 2002).

The relative performance of these procedures is then measured in a controlled lab experiment for several preference profiles. *The key finding from this*

[3]The Supplementary Appendix contains a list of agencies that use this procedure.

[4]Anbarci (2006) introduced this variant because AS generates a SCR that is sensitive to the removal of Pareto inferior arbitrators from the list. The SCR associated to VAOV does not suffer from that defect.

analysis is that two sequential procedures that are not used in practice, namely SL and VAOV, dominate the commonly used VR mechanism. The AS and VR are not comparable as there are preference profiles for which the former dominates the latter, and others for which the comparison is reversed. Similarly, there is no definite ranking between SL and VAOV.

The paper unfolds as follows. After discussing the related literature, the empirical context is presented in Section 2. Section 3 contains theoretical results (proofs are relegated to the appendix). Experimental design and data analysis are available in Section 4. The concluding section summarizes our findings.

Related Literature

The most closely related paper is Bloom and Cavanagh (1986a), who analyze the selection of arbitrators using data on arbitration cases from the New Jersey Public Employment Relations Commission (PERC) during 1980. Data are based upon the simultaneous veto-rank scheme described in the Introduction (with $n = 7$). Their analysis first examines the degree of overlap between rankings in order to shed light on the similarity of preferences. They show some, but not complete, overlap in rankings, and, under the assumption of sincere rankings, conclude that there is some, but not complete, overlap in preferences. We will reach the same conclusion, but without assuming that parties are truthful in their reports.

Their second analysis uses rankings and characteristics of arbitrators to measure the degree to which certain characteristics are valued by the different parties. They find, for example, that employers rank economists more highly than unions do. Under an assumption of sincere rankings, one can conclude that employers have a relative taste for economists and that unions have a distaste for economists. The assumption of sincere rankings is debatable though, and indeed we present theoretical and experimental evidence that it does not hold. Bloom and Cavanagh try to address this issue by fitting their model under the weaker assumption strategic players always rank their most preferred alternative first but may strategize on other dimensions of their report. They observe that their preference parameter estimates do not vary much when us-

ing only the first choice data, and conclude from it that there is no evidence of strategic play. A key limitation of this test involves the breakdown of the assumption that strategic players always rank their most preferred alternative first. It is straightforward to generate counter-examples to this: if the union vetoes the first choice of the employer, the employer may choose to not rank their most preferred alternative first as this is "wasting" the first ranking. Indeed, we present evidence from one of our experiments below documenting that a substantial fraction of players do not rank first their most preferred alternative when it is not viable, in the sense of being the worst for their opponents.

In an unpublished working paper, Bloom and Cavanagh (1986b) discuss some theoretical properties of the VR and AS mechanisms. They show that the former has non-truthful and inefficient equilibria, while the equilibria of the latter are all efficient. They also show that if the parties held uniform priors over all the possible strict rankings of arbitrators, then being truthful is an efficient Bayesian Nash equilibrium in both mechanisms.[5] Our focus, however, is on the implementation-theoretic view of arbitrator selection. In particular, we show that a large class of SCRs with appealing properties is impossible to implement, while alternative SCRs are implementable by "natural" mechanisms.

More generally, this present paper is related to a literature on matching, where economists have identified market failures and proposed new mechanisms that solve these failures. Several of these mechanisms, similarly to the veto-rank scheme used in selecting arbitrators, involve participants submitting rank-ordered preferences. Examples include mechanisms for matching residents to hospitals and students to elementary schools (see Roth (1984, 2007) and Abdulkadiroglu, Pathak, and Roth (2005a and 2005b)). This literature has focused on implementing strategy-proof mechanisms using variants of the Gale-Shapley deferred acceptance algorithm or the top-trading cycle

[5]One complication that arises when analyzing Bayesian Nash equilibria, especially in the veto-rank game, is that one needs to make assumptions about each player's belief about his opponent's preferences over *lotteries*.This concern, however, is not discussed in their paper.

mechanism. In the context of the selection of arbitrators, there is no deterministic mechanism that is strategy-proof and instead we propose and analyze alternative sequential mechanisms.

Given our focus on whether participant ranks and vetoes are sincere or strategic, this paper is also related to a literature on strategic voting, which can take many forms. In an experimental setting with three candidates and plurality rule, Forsythe, Myerson, Rietz, and Weber (1993 and 1996) find substantial evidence that voters are strategic in the sense of not voting for their most preferred candidate when this candidate has little chance of winning. Focusing on the case of bundled elections, Degan and Merlo (2007) find little evidence that voters are strategic in the sense that they might account for the fact that policy outcomes may depend upon both the Congress and the President. In a model with incomplete information, Kawai and Watanabe (2010) estimate that a large fraction of voters in Japanese elections are strategic in the sense of conditioning on the state of the world where they are pivotal.

2 Are Preferences Perfectly Opposed?

The question of designing an appropriate selection procedure is relevant only if preferences are *not* strictly opposed. Otherwise, the VR mechanism is a simple zero-sum game, and all its Nash equilibria are reasonable. So is it the case that parties have strictly opposed preferences? As explained in the Introduction, there are reasons to believe otherwise. Beyond intuition, this section presents an empirical test showing that preferences are not always opposed.

We use information from the New Jersey Public Employment Relations Commission (PERC). During the years 1985 to 1996, employers and unions were provided a menu of seven arbitrators and were asked to veto three arbitrators and to rank the remaining four.[6] The arbitrator with the lowest combined rank among those that were not vetoed by either party was then chosen as the arbitrator for the case. This mechanism thus corresponds to the

[6] After 1996, the an arbitrator is randomly selected by a computer from the list of approved arbitrators.

VR mechanism described in the Introduction, with $n = 7$.

Data on rankings by employers and unions cover the years 1985 to 1996. Variables in this dataset include the year of the case, the names of the two parties (employer and union), the menu of arbitrators (including first and last name), the rankings of each party, and the name of the arbitrator chosen by the procedure. Employers in these data are local governments within the state of New Jersey.[7] We drop a number of cases with inconsistent or incomplete data.[8] After deleting these observations, we are left with 750 cases with complete rankings by employers and employees and in which the chosen arbitrator followed the rankings submitted by the two parties. Given that the menu includes seven arbitrators, we thus have 5,250 arbitrator choices and 10,500 unique ranks, one for the employer and one for the union.

When preferences are perfectly opposed, truthful behavior is a Nash equilibrium. Moreover, truthful play may be considered the focal equilibrium in this case. Thus, if preferences were perfectly opposed, then there should be no overlap in terms of either the rankings or the vetoes.

	U1	**U2**	**U3**	**U4**	**U veto**	**total**
E1	2.38%	1.83%	2.19%	1.70%	6.19%	14.29%
E2	2.13%	2.34%	2.13%	1.83%	5.85%	14.29%
E3	2.13%	2.00%	1.89%	1.94%	6.32%	14.29%
E4	1.62%	2.11%	1.90%	2.25%	6.40%	14.29%
E veto	6.02%	6.00%	6.17%	6.57%	18.10%	42.86%
total	**14.29%**	**14.29%**	**14.29%**	**14.29%**	**42.86%**	**100%**

Table 1: Distribution of Employer (E) and Union (U) Rankings of Arbitrators

[7]These include municipalities, such as the city of Trenton, agencies within municipal governments, such as corrections in Middlesex County, and agencies within the state government, such as the New Jersey State Police. Unions represented are then public sector unions within the relevant government or government agency.

[8]Examples of cases with inconsistent or incomplete data include some cases in which the arbitrator chosen does not reflect the mechanism described above, cases in which one or both of the two parties did not submit a ranking, and cases in which parties submitted rankings but did not follow the request to veto three options and rank the remaining four.

As shown in Table 1, however, we see little evidence of perfect opposition in terms of rankings.[9] For example, conditional on an arbitrator being ranked first by the union, this arbitrator is also ranked first by the employer in 17 percent of cases, ranked second in 15 percent of cases, ranked third in 15 percent of cases, ranked fourth in 11 percent of cases, and vetoed in 42 percent of cases. The nearly uniform distribution across these categories is thus inconsistent with perfect opposition.

One limitation of this test involves multiplicity of equilibria. In particular, when preferences are perfectly opposed, parties always veto their bottom $(n-1)/2$ ranked options in equilibrium but any ranking of the remaining $(n+1)/2$ options constitutes a Nash equilibrium. Indeed, we will see in our experimental analysis to follow that subject pairs are truthful in only 25% of cases when preferences are completely opposed, and this is due mainly to deviations from sincerity on the ranking of the non-vetoed options.

To address this issue, we next develop a more robust test based upon the fact that, as noted above, there should be no overlap in vetoes when preferences are perfectly opposed. Based upon an analysis of vetoes, however, we find a substantial degree of overlap in vetoes. In particular, in 50% of cases there is one common veto, in 34% of cases there are two common vetoes, and in 3% of cases there are three common vetoes. That is, in 87% of the cases the parties' vetoes overlap, and there is no overlap in only 13% of cases. Taken together, these tests find little evidence that preferences are perfectly opposed.

3 Theoretical Motivation

Two parties, $i = 1, 2$, face a finite set A of $n \geq 4$ candidates that an agency proposes as potential arbitrators. We assume that n is odd, as this is the scenario favored by arbitration agencies and studied in our experimental analysis (all the results in this section can be extended to the case where n is even). $\mathcal{P}$ denotes the set of strict preference relations $\succ$ on A. Most disputes resolved through arbitration occur between parties that have a long-term relationship

[9] We can also use this evidence to argue that preferences are not perfectly aligned.

(e.g., unions and managements). In addition, arbitration agencies provide both parties with the same detailed resumés of the potential arbitrators. Hence it is not unreasonable to assume that the parties' ordinal preferences are commonly known among them (put differently, we consider implementation under "complete information").

Definition 1 *A* social choice rule *(SCR) is a correspondence* $f : \mathcal{P} \times \mathcal{P} \to A$ *such that* $f(\succ_1, \succ_2)$ *is a non-empty subset of* A*, for each* $(\succ_1, \succ_2) \in \mathcal{P} \times \mathcal{P}$*.*

Definition 2 *A SCR* f *is* partially implementable *if there exists a* mechanism (S_1, S_2, μ)*, where* S_i *is* i*'s strategy set and* $\mu : S_1 \times S_2 \to A$ *is the outcome function, such that, for each* $(\succ_1, \succ_2) \in \mathcal{P} \times \mathcal{P}$*, the set of pure-strategy Nash equilibrium outcomes associated to the strategic-form game* $(S_1, S_2, \mu, \succ_1, \succ_2)$ *is non-empty and a subset of* $f(\succ_1, \succ_2)$*.*

Notice that the veto-rank procedure discussed in the Introduction does not qualify as a mechanism in this sense, because the outcome function delivers a lottery in some circumstances. Considering lotteries, and thinking about how parties behave when facing such uncertainty, leads us to consider risk preferences. Let $\mathcal{U}$ be the set of strict Bernoulli functions (the defining ingredient of von Neumann-Morgenstern preferences). A typical element u of $\mathcal{U}$ is thus simply a function $u : A \to \mathbb{R}$, with $u(a) \neq u(a')$ whenever $a \neq a'$, and preferences between lotteries over A are derived by computing expected utility with respect to u. It is less plausible to think that there is complete information regarding these Bernoulli functions, but our analysis is robust against that assumption in that our sole objective when considering lotteries is to show that strong negative results hold *even if* there was complete information in that regard.

Definition 3 *A* random social choice function *(RSCF) is a function* $\psi : \mathcal{U} \times \mathcal{U} \to \Delta(A)$ *that associates a lottery to each pair of strict Bernoulli functions.*

Definition 4 *The RSCF* ψ *is* implementable *if there exists a* random mechanism (S_1, S_2, μ)*, where* S_i *is* i*'s strategy set and* $\mu : S_1 \times S_2 \to \Delta(A)$ *is the*

outcome function, such that, for each $(u_1, u_2) \in \mathcal{U} \times \mathcal{U}$*, any pure-strategy Nash equilibrium outcomes associated to the strategic-form game* $(S_1, S_2, \mu, u_1, u_2)$ *coincides with* $\psi(u_1, u_2)$*.*

Procedure 1 (*Veto-Rank*) (VR) The veto-rank procedure provides an example of random mechanism. Both parties ($i = 1, 2$) simultaneously choose a pair (V_i, r_i), where V_i is a set of vetoed options that contains $\frac{n-1}{2}$ elements from A, and r_i is a scoring rule that assigns to every element in $A \backslash V_i$ an integer from zero to $n - k - 1$ such that no two elements are assigned the same score. The outcome is determined as follows. If $A \backslash (V_1 \cup V_2)$ is a singleton, then this arbitrator is chosen. Otherwise, an element in $A \backslash (V_1 \cup V_2)$, is selected by maximizing the sum of scores, $r_1(\cdot) + r_2(\cdot)$, with ties being broken via a uniform lottery.

For each $a \in A$ and each $u \in \mathcal{U}$, let $\sigma(a, u) = \#\{a' \in A | u(a') < u(a)\}$. The veto-rank procedure is played truthfully if, for each $(u_1, u_2) \in \mathcal{U} \times \mathcal{U}$ and both $i = 1, 2$, the set V_i contains the $\frac{n-1}{2}$ worst elements according to u_i, and $r_i(a) = \sigma(a, u_i) - \frac{n-1}{2}$, for each element $a \in A \setminus V_i$. This generates the following natural RSCFs. For each $(u_1, u_2) \in \mathcal{U} \times \mathcal{U}$, $\psi_{VR}(u_1, u_2)$ will denote the uniform lottery defined over

$$\arg\max_{a \in X(u_1, u_2)} (\sigma(a, u_1) + \sigma(a, u_2))$$

where

$$X(u_1, u_2) = \{a \in A | \sigma(a, u_i) \geq \frac{n-1}{2}, \text{ for } i = 1, 2\}.$$

The support of ψ_{VR} also defines a natural SCR: for each $(\succ_1, \succ_2)$,

$$f_{VR}(\succ_1, \succ_2) := support(\psi_{VR}(u_1, u_2)),$$

where u_i is any[10] strict Bernoulli function that is consistent with $\succ_i$ over A.

Preliminary Observations. *The VR procedure has the following properties.*

a) (non-truthfulness) *Truthtelling is not a Nash equilibrium for some preference profiles, and for every preference profile there is a non-truthtelling Nash*

[10]Notice indeed that ψ_{VR} varies only with the ordinal information encoded in the Bernoulli functions.

equilibrium.

b) (undesirable equilibria) *There are preference profiles for which the mechanism induces Nash equilibrium outcomes that are not selected by* f_{VR}.

c) (risk of miscoordination) *There are preference profiles for which there exists a pair of equilibria,* $s = (s_1, s_2)$ *and* $s' = (s'_1, s'_2)$, *such that if both players coordinate on either* s *or* s' *the resulting outcome is in* $f_{VR;}$, *but if one player follows* s *and another follows* s', *the resulting outcome is Pareto inefficient.*

These preliminary observations raise the questions of whether there exists another normal-form mechanism that implements the RSCF ψ_{VR}, or that partially implements the SCR f_{VR}. We believe that the main reason why arbitration agencies aim to implement f_{VR} is that all the outcomes that emerge with positive probabilities satisfy the following two properties. A RSCF ψ is *Pareto efficient* if, for each (u_1, u_2) and each x in the support of $\psi(u_1, u_2)$, it is impossible to find $a \in A$ such that $u_i(a) > u_i(x)$ for both $i \in \{1, 2\}$. It passes the *minimal satisfaction test* (MST) if $\sigma(x, u_i) \geq \frac{n-1}{2}$ for each $i \in \{1, 2\}$, each $u \in \mathcal{U} \times \mathcal{U}$, and each x in the support of $\psi(u_1, u_2)$. Similar definitions also apply to SCRs. The SCR f_{VR} and the RSCF ψ_{VR} are both Pareto efficient and both pass the minimal satisfaction test. However, the next proposition shows that any SCR (or RSCF) that satisfies these two properties is not implementable.

Proposition 1 *The following three statements hold.*

a) There is no SCR that is partially implementable, Pareto efficient, and that passes the MST.

b) There is no RSCF that is implementable, Pareto efficient, and that passes the MST.

c) In particular, ψ_{VR} *is not implementable, and* f_{VR} *is not partially implementable.*

Proposition 1 implies that the only hope of implementing f_{VR} is to use extensive-form mechanisms. Because our goal is to consider mechanisms that are potentially applicable, we focus on finite extensive-form mechanisms of

perfect information, which are thus solvable by backward induction. In addition, we focus on a notion of implementability that combines ideas of partial implementability of SCRs and implementability of RSCFs.

Definition 5 *A SCR f is* fully implementable by backward induction *if there exists a two-player extensive-form mechanism of perfect information such that, for each $(\succ_1, \succ_2) \in \mathcal{P} \times \mathcal{P}$, $f(\succ_1, \succ_2)$ coincides with the union of the two subgame perfect equilibrium outcomes associated with the two extensive-form games obtained when assigning either the first or the second party to the role of the first player.*[11]

A fully implementable SCR naturally leads to an RSCF by tossing a fair coin to randomly select an element of the SCR. This associated RSCF is clearly implementable by backward induction, via the extensive-form where chance decides in a first move who will assume the role of the first player. Implementability, efficiency and the MST become compatible when considering backward induction in this larger class of mechanisms. However, even in this larger class, f_{VR} is impossible to implement.

Proposition 2 *There is no single-valued selection of f_{VR} that is implementable by backward induction.*

It follows that in order to guarantee desirable outcomes, we must consider alternative (desirable) SCRs. We, therefore, introduce two sequential mechanisms whose subgame perfect equilibria have been previously characterized by Anbarci (2006). The first procedure, the *Alternate-Strike*, is being used by some agencies. The second procedure, *Voting by Alternating Offers and Vetoes*, was proposed by Anbarci to derive a SCR that is immune to changes when removing a Pareto inferior arbitrators from the list. As far as we know, this second sequential mechanism is not used in practice.

Procedure 2 (*Alternate Strikes*) (AS) Both parties take turns removing an arbitrator from the set A until the last remaining arbitrator is chosen.

[11] Preferences being strict, backward induction always leads to a unique outcome in each extensive-form game of perfect information.

Procedure 3 (*Voting by Alternate Offers and Vetoes*) (VAOV) One party, call it player 1, proposes an option $a \in A$ to the other party, call it player 2, who may either accept or reject. If 2 accepts, a is chosen; otherwise, a is removed, and player 2 proposes to 1 an option $b \in A\backslash\{a\}$, which player 1 may either accept or reject. The game continues until one of the options is accepted or until only one option remains, which is then chosen.

Proposition 3 *(Anbarci, 2006) The following two statements hold.*

a) The AS procedure fully implements the SCR f_{AS}, which can be computed inductively as follows. Let $A_0(\succ) = A$, and, for any integer $t \geq 1$, let

$$A_t(\succ) = A_{t-1}(\succ)\backslash\{w_1(A_{t-1}(\succ), \succ), w_2(A_{t-1}(\succ), \succ)\},$$

where $w_i(A_{t-1}, \succ)$ is i's least preferred arbitrator within A_{t-1}. Then $f_{AS}(\succ) = A_{t^*-1}(\succ)$, where t^* is the smallest integer such that $A_{t^*} = \emptyset$.

b) The VAOV procedure fully implements the SCR f_{VAOV}, which can be computed as follows:

$$f_{VAOV}(\succ_1, \succ_2) = \arg\max_{a \in A} \min_{i=1,2} \sigma(\succ_i, a),$$

where $\sigma(\succ_i, a) = \#\{a' \in A | a \succ_i a'\}$.

It is easy to check that f_{AS} and f_{VAOV} are both Pareto efficient and pass the MST. Also, f_{VAOV} has already been studied previously as a reasonable SCR independently of its implementability, see Sprumont's (1993) "Rawlsian arbitration rule," Hurwicz and Sertel's (1997) "Kant-Rawls social compromise," Brams and Kilgour's (2001) "fallback bargaining," and Kibris and Sertel's (2007) "unanimity compromise."

Notice that the SCRs f_{VAOV} and f_{VR} share the common feature of using scores based on the two parties ordinal rankings, a tradition that goes back at least to the 18^{th} century with Borda. While f_{VR} uses these scores in a utilitarian tradition, summing them up, f_{VAOV} uses them in an egalitarian tradition, aiming at maximizing the welfare index of the worse-off party. Vetoes were needed for the SCR f_{VR} to pass the MST. Applying the egalitarian criterion

instead guarantees that the resulting SCR passes that test without the need to resort to vetoes.

It is well documented that backward induction does not systematically prevail when the game has multiple stages (see e.g. Binmore et al. (2002) and Levitt, List and Sadoff (2011)). There are thus reasons to focus on short extensive forms. We now investigate the simplest possible such mechanisms. Formally, a *two-stage mechanism* is composed of a finite set of actions A_1 for the first mover (the identity of which can be chosen by tossing a fair coin) a function A_2 that determines a finite set of actions for the second mover (the other party) as a function of the first mover's choice, and an outcome function o that selects an element in A for each pair (a_1, a_2) such that $a_1 \in A_1$ and $a_2 \in A_2(a_1)$.

Backward induction leads to an optimal strategy for the second mover: $a_2^*(a_1, \succ_2) \in A_2(a_1)$, for each $a_1 \in A_1$ and each $\succ_2 \in \mathcal{P}$, such that $o(a_1, a_2^*(a_1, \succ_2))$ is optimal according to $\succ_2$ within $\{o(a_1, a_2) | a_2 \in A_2(a_1)\}$. Then the optimal strategy for the first mover is given by $a_1^*(\succ)$ where $o(a_1^*(\succ), a_2^*(a_1^*(\succ), \succ_2))$ is optimal according to $\succ_1$ within $\{o(a_1, a_2^*(a_1, \succ_2)) | a_1 \in A_1\}$, for each $\succ \in \mathcal{P} \times \mathcal{P}$. Let $o^* : \mathcal{P} \times \mathcal{P} \to A$ be the outcome of the two-stage mechanism when played by backward induction: $o^*(\succ) = o(a_1^*(\succ), a_2^*(a_1^*(\succ), \succ_2))$, for each $\succ \in \mathcal{P} \times \mathcal{P}$. The function o^* is the SCR implemented by the two-stage mechanism (A_1, A_2, o).

Proposition 4 *There exists a unique single-valued SCR o^* that is Pareto efficient, passes the MST, and can be implemented by backward induction via a two-stage mechanism. It is computed as follows:*

$$o^*(\succ) = \arg\max_{\succ_1} \{a \in A | \#\{b \in A | a \succ_2 b\} \geq \frac{n-1}{2}\}.$$

In addition, o^* is implementable via the following two-stage mechanism:

Procedure 4 *(Shortlisting)* (SL) The party that has been selected to be the first mover chooses a subset containing $\frac{n+1}{2}$ elements of A, and the other party subsequently picks an arbitrator out of that subset $A_2(a_1) = a_1$.

While truthtelling (i.e., selecting his top $(n+1)/2$ elements) is not always an equilibrium strategy for the first mover, an equilibrium strategy (or a best

response to the belief that his opponent is rational) can be derived using the following simple algorithm. In the first step, player 1 checks if there is a set S of $(n-1)/2$ elements that player 2 ranks below 1's top choice, a. If so, 1 chooses $\{a\} \cup S$. Otherwise, he goes to the next step. In the second, step player 1 checks if there is a set T of $(n-1)/2$ elements that 2 ranks below 1's second-best choice, b. If so, 1 chooses $\{b\} \cup T$. Player 1 continues in this fashion until the algorithm terminates at or before 1's median choice.

4 Experimental Analysis

The Preliminary Observations in the previous section highlighted a number of theoretical concerns with using the VR mechanism. Of course, none of these issues would be of any concern if the participants do not behave strategically as the theory assumes. In particular, the mechanism would attain desirable outcomes if parties naïvely delete worse options and truthfully report their ranking for the remaining arbitrators.

In contrast, the sequential mechanisms studied here induce equilibrium outcomes that implement closely related SCRs/RSCFs and satisfy the basic properties of Pareto efficiency and minimal satisfaction. This result relies on the participants' abilities to perform backward induction. However, a number of studies in the experimental literature suggest that most subjects find it difficult to perform backward induction, and often fail to carry it out.

Thus, in order to evaluate the performance of the mechanisms defined in the previous section, it is important to understand how people actually behave when implementing these mechanisms. Since this requires us to know the parties' true preferences, it is difficult to draw conclusions from empirical data. We, therefore, conducted a series of computerized laboratory experiments to test all four mechanisms in a controlled environment.

4.1 Design

The experiments were conducted at NYU's Center for Experimental Social Science. A total of 304 subjects from the undergraduate student population

participated.

In each treatment, an even number of subjects was presented with a set of five alternatives, $A = \{a, b, c, d, e\}$, and were randomly matched to play one of the mechanisms on this set of options. Each treatment consisted of 40 rounds, which were divided into four "blocks" of ten rounds. In each of these blocks, subjects had the same preference relation over the five options, but these preferences changed from one block to another (i.e., in total there are four distinct preference profiles). Preferences over A are induced by assigning each of the options a distinct monetary value in the set $\{\$1.00, \$0.75, \$0.50, \$0.25, \$0.00\}$. The four preference profiles were as follows

Pf1		**Pf2**		**Pf3**		**Pf4**		
Pl. 1	Pl. 2	Pl. 1	Pl. 2	Pl. 1	Pl. 2	Pl. 1	Pl. 2	Payment
a	e	a	b	a	c	a	e	\$1.00
b	d	b	a	b	b	b	c	\$0.75
c	c	c	c	c	a	c	a	\$0.50
d	b	d	d	d	d	d	b	\$0.25
e	a	e	e	e	e	e	d	\$0

Table 2: Four Preference Profiles Tested in the Experiment

The first profile, Pf_1, consists of completely opposed rankings. The second profile, Pf_2, represents partial conflict of interest involving only the top two options. This is a case where truthtelling does not form a Nash equilibrium and where there is a risk of bad outcome due to miscoordination (see proof of Preliminary Observations a) and b) in the previous section). The third profile, Pf_3 displays a similar partial conflict of interest at the top, but this time with the addition of a focal compromise (b). The fourth profile, Pf_4, captures cases where the veto-rank mechanism admits (undominated) Nash equilibria whose outcome do not belong to the veto-rank SCR (see Preliminary Observation b) in the previous section).

There were four treatments, each corresponding to one of the mechanisms we discussed. There were 70 participants in the first treatment, 74 in the second, 72 in the third and 88 in the fourth. For each mechanism and each

preference profile, we have characterized the set of pure-strategy equilibria.[12] For each treatment we ran four sessions, where in each session the four induced preference profiles appear in a different order. The four orders were: $Pf_1 - Pf_2 - Pf_3 - Pf_4$, $Pf_4 - Pf_3 - Pf_2 - Pf_1$, $Pf_1 - Pf_3 - Pf_2 - Pf_4$, and $Pf_4 - Pf_2 - Pf_3 - Pf_1$. Hence, each profile was played (by a different group of subjects) at two different stages in the experiment: an "early" stage (the first ten rounds for Pf_1 and Pf_4 and the second block of ten rounds for Pf_2 and Pf_3) and a "late" stage (the last ten rounds for Pf_1 and Pf_4 and the third block of ten rounds for Pf_2 and Pf_3). This allows us to examine whether there was a learning "spillover" from one profile to another, see below.

Subjects were paid the sum of their earnings across the 40 rounds in addition to a show-up fee of \$10. The Supplementary Appendix contains the instructions to one of the treatments (instructions to the other treatments were similar and are available from the authors upon request). After subjects read the instructions they were presented with a short quiz that tested their understanding of the game. When the subjects finished answering the quiz, they were presented with the correct answers. The instructions in the Supplementary Appendix also include the quiz that followed them.

4.2 Strategic Behavior and Outcomes in VR

As explained earlier, the veto-rank mechanism delivers appealing outcomes when participants are truthful, with both participants vetoing their bottom two options and ranking the remaining three in accordance to their preferences. Yet there are theoretical reasons to believe that participants would not be truthful, and strategize instead. Do participants in the VR procedure tend to be truthful?

Our data on actual arbitration cases from the state of New Jersey provide suggestive evidence for strategic behavior. In particular, our data contains

[12] It is straightforward to verify whether a pair of actions constitute an equilibrium in the veto-rank and the shortlisting games. The characterization for AS and VOAC is more involved and is available from the authors upon request. Equilibrium strategies for the shortlisting scheme were described in the previous section.

249 instances in which the same employer had the same two arbitrators in his choice set in two different arbitration cases, and neither arbitrator was selected in these two cases, nor in any case during the period between them. Under the assumption that an employer's relative ranking of an arbitrator can change only as a result of direct experience with that arbitrator, a truthful employer should treat the two arbitrators in the same way in both cases. In roughly-one third of the 249 observations, however, an employer reverses his ranking of the two arbitrators. Further details of these tests are available in the Supplementary Appendix.

Due to the strong stability assumption underlying this empirical test, we next investigate the issue of truthful behavior in the lab.

Result 1 *A majority of subjects in the experiment are not truthful. Those who do not play truthfully appear to follow some strategic motives instead of playing randomly.*

SUPPORT: The next table reports the observed percentage of subjects who played truthfully, as a function of the preference profile.

	Pf1	**Pf2**	**Pf3**	**Pf4**
% Truthful	50%	43%	32%	26%

Table 3: Percentage of Subjects Who Played Truthfully

These numbers constitute upper bounds on the percentage of "naïve" participants who did not strategize. This is because truthful behavior may be a best response against the other party's strategy. In addition, for some preference profiles (e.g. Pf1), there is a Nash equilibrium in which both parties are truthful.

We next turn to the question of whether participants, who strayed from truthful behavior, did so for strategic reasons. One way of addressing this question is to compute the percentage of subjects whose choice of strategy is part of some Nash equilibrium. However, such a test is not very informative, as far too many strategies exhibit this property in the VR mechanism. We

therefore take a different approach and adopt the framework of k-level reasoning (see the survey in Crawford et al. 2010). The natural candidate for level zero (non-strategic) behavior is being truthful. Level 1 would then constitute a best response against truthful behavior. Table 4 depicts the percentages of Level 1 choices in the data.

	Pf1	**Pf2**	**Pf3**	**Pf4**
% Level 1 Among Non-Truthful	63%	69%	28%	47%

Table 4: Percentage of Non-Truthful Subjects Who Played Level 1

The p-value for getting a percentage higher than or equal to these under the assumption that non-truthful subjects play randomly is less than 0.0001 for each of the four preference profiles.[13] □

Notice that a good outcome is guaranteed in the veto-rank mechanism if both participants play truthfully. If x% the participants play truthfully, as reported in Table 3, then on average only (x^2)% of the matched pairs have both participants play truthfully, which is even significantly lower.

The fact that subjects are not truthful in itself is not problematic, provided that final outcomes are desirable. Unfortunately, the theory section provides arguments that strategic considerations are likely to lead to undesirable outcomes. Even though subjects' actions naturally do not exactly match theoretical predictions, data from the experiment does confirm the insight derived from the theory.

Result 2 *A significant proportion of observed outcomes for the Veto-Rank procedure are inefficient and/or fail the MST.*

SUPPORT: The following table shows the percentage of observed outcomes which are inefficient and/or fail the MST, as a function of the preference profile.

[13]It is tempting to push the strategic analysis further, and investigate for instance what part of the data can be explained by adding higher levels of reasoning. At least 97% of observed choices can be explained by levels 0, 1 and 2 in all four preference profiles. Yet this is not very informative because each strategy admits many best responses. In particular, between 75 and 90% (depending on the preference profile) of strategies belong to one of these three levels. Level 1, on the other hand, contains only between 10 and 25% of all strategies, explaining why we focus on this particular category.

	Pf1	Pf2	Pf3	Pf4
% Inefficient	0%	9%	12%	19%
% Failing MST	27%	3%	12%	21%
% Inefficient or Failing MST	27%	9%	12%	21%

Table 5: Percentage of Outcomes that Are Inefficient and/or Fail the MST

(Each outcome is Pareto efficient with $Pf1$ since preferences are completely opposed, hence the 0% entry in that case.) □

4.3 Comparing Outcomes Across Procedures

According to the theory, each of the three sequential procedures should dominate VR, according to our two criteria of Pareto efficiency and minimal satisfaction. However, it is not clear that individuals actually behave according to the theory. There is evidence, for instance, that people have difficult performing backward induction for multiple rounds (see e.g Binmore et al., 2002). Judging observed outcomes in terms of efficiency and/or the MST, the following result indicates which of our three dynamic procedures unequivocally dominate VR: SL and VAOV do, but not AS. There is no definite comparison between the two procedures that dominate VR.

Result 3 *Both SL and VAOV dominate VR. There are preference profiles for which AS outperforms VR, and others where the comparison is reversed. Hence AS and VR cannot be ranked. VAOV and SL cannot be ranked either.*

Support: The next table shows the percentage of matches whose outcome failed the efficiency criterion and/or the MST, as a function of the preference profile.

	Inefficient			Fails MST				Inefficient or Fails MST			
	Pf2	**Pf3**	**Pf4**	**Pf1**	**Pf2**	**Pf3**	**Pf4**	**Pf1**	**Pf2**	**Pf3**	**Pf4**
VR	9%	12%	19%	27%	3%	12%	21%	27%	9%	12%	21%
AS	17%	4%	14%	16%	3%	4%	17%	16%	17%	4%	17%
VAOV	5%	2%	12%	13%	2%	2%	19%	13%	5%	2%	19%
SL	3%	11%	7%	18%	1%	11%	10%	18%	3%	11%	10%

Table 6: Percentage of Outcomes That Are Inefficient and/or Fail MST

The next table reads as follows. We write $P_1 > P_2$ if the observed proportion of outcomes that fail the criterion under study (efficiency and/or MST) is smaller when the procedure P_1 is implemented than when procedure P_2 is implemented. If the difference in proportions is statistically significant (bilateral test at 5% level[14]), then we write $P_1 \succ^* P_2$. Notice that the binary relation, "P_1 dominates P_2 in a statistical significant way", is not transitive. Nevertheless, our data has the property that for any triplet of procedures, P_1, P_2 and P_3, it is true that $P_1 \succ^* P_3$ whenever $P_1 > P2$ and $P_2 \succ^* P_3$. All the p-values are available in the Supplementary Appendix.

	Pf1	**Pf2**
Eff	NA	$SL > VAOV \succ^* VR \succ^* AS$
MST	$VAOV > AS > SL \succ^* VR$	$SL > VAOV > AS = VR$
Eff & MST	$VAOV > AS > SL \succ^* VR$	$SL > VAOV \succ^* VR \succ^* AS$

Table 7a: Percentage of Outcomes that Are Inefficient and/or Fail the MST

	Pf3	**Pf4**
Eff	$VAOV > AS \succ^* SL > VR$	$SL \succ^* VAOV > AS \succ^* VR$
MST	$VAOV > AS \succ^* SL > VR$	$SL \succ^* AS > VAOV > VR$
Eff & MST	$VAOV > AS \succ^* SL > VR$	$SL \succ^* AS > VAOV > VR$

Table 7b: Percentage of Outcomes that Are Inefficient and/or Fail the MST

Tables 7a-7b show that VAOV dominates VR, with a higher proportion of desirable outcomes in all preference profiles. This difference in proportions is

[14]Equality is also rejected with much tighter confidence levels in most cases.

statistically significant for both criteria and all preference profiles, except for MST in the profile Pf4. Similarly, SL dominates VR with a higher proportion of desirable outcomes in all preference profiles. Apart for Pf3, this difference in proportions is also statistically significant for both criteria and all preference profiles. AS dominates VR for Pf1, Pf3 and Pf4 in a way that is statistically significant, but VR dominates AS for Pf2 in a way that is statistically significant. Similarly, SL and VAOV are not systematically comparable, as VAOV dominates SL for Pf3 in a way that is statistically significant, but SL dominates VAOV for Pf4 in a way that is statistically significant. □

4.4 Social Preferences

If subjects care only about their own material payoffs and follow equilibrium play, then observed outcomes for the three sequential procedures would systematically meet our two criteria of efficiency and minimal satisfaction. We anticipated the risk that subjects may have difficulties in performing many rounds of backward induction. Under that hypothesis, one would have expected SL to systematically perform better than any of the other procedures. Yet, we see from Table 7 that this conclusion is contradicted for Pf3. Similarly, we anticipated that AS would have outperformed VR, or at least performed in a comparable way to VAOV, given that backward induction seems to involve a similar depth of reasoning in both cases. Yet we see again from Table 7 that this conclusion is contradicted for Pf2. These unanticipated results suggest that participants are motivated by social preferences in the sense that they care not only about their own monetary payoffs, but also on how they are treated by others.

Most of implementation theory ignores the potential impact of social preferences whereby the players' preferences may be endogenously determined by the mechanism (recent exceptions include Bowles and Hwang (2011) and Bierbrauer and Netzer (2012)). Our next result provides evidence that seemingly similar mechanisms induce different other-regarding preferences in participants.

Result 4 *Social preferences appear to drive the choice of a significant number of subjects in AS with Pf2, and in SL with Pf3. The VAOV procedure shares some similarity with AS. Yet the fact that subjects make proposals that can be rejected, instead of directly deleting options from the list, appear to make social preferences significantly less salient in VAOV than in AS.*

SUPPORT: Notice that, given the conflicting interests in Pf3, option b appears to strike a reasonable compromise. However, the first mover can achieve a, a better outcome for him, by excluding b from the shortlist. A significant proportion (19%) of second movers who received the shortlist with their bottom three options (as should be in the unique subgame perfect equilibrium), sacrifice their own material interest by picking an option that punishes the first-mover whose action may have appeared greedy (see Charness and Rabin (2002)). This explains all but one of the inefficient outcomes we observed for SL with Pf3. A significant proportion of first movers (42%) also departed from their optimal selfish strategy by including b in the shortlist. Presumably this was done either to appear fair, or to avoid retaliation, if they think their opponent may be offended by a shortlist that appears more greedy. This explains why many b outcomes are observed even though they are not supported by backward induction.

The action of the first-mover in any subgame perfect equilibrium for AS with Pf2 is quite simple. He must veto his opponent's top choice, in order to get his own top choice (the last remaining option that is Pareto dominant). Hence, one would expect the AS procedure to deliver only a or b as outcomes (depending on who moves first), in which case AS would perform well in terms of efficiency and MST. Instead, we observe a significant (15%) number of c's, and a few d's and e's. Again, these may be explained by social preferences: 18% of the cases where a subject vetoed his opponent's top choices while both a and b were still available, were followed by the opponent crossing that subject's optimal choice (which is suboptimal from a selfish perspective). Subjects seemed to be aware of this risk, as 33% of first-movers did *not* cross their opponent's top choice.

We next turn to examine the subjects' behavior in VAOV under the profile

Pf2. The subgame perfect equilibrium outcome is the same as for AS when subjects care only about their own payoff, with the first-mover getting his most preferred option. A play path with either one of the following two properties may be viewed as an instance of social preference: (*i*) the first-mover proposes his top pick which gets rejected, and he retaliates by rejecting the second mover's top pick when offered, or (*ii*) the first-mover proposes his opponent's top choice (presumably because he wants to avoid any risk of retaliation). Even with this encompassing definition of social preferences, only 2% of matched pairs fall in this category. Thus, even though having a subject i reject a proposal to implement j's top alternative in VAOV is equivalent (in terms of outcomes) to i eliminating j's top choice in AS, it may not be perceived the same way. One reason for this may be that it seems only fair that a greedy proposal by one of the players is not accepted. □

We conjecture that the role of social preferences may be diminished when larger stakes are involved with experienced or professional players, in which case the appeal of the AS and SL procedures would further increase. One possible way to test this in the future would be to run a similar experiments with arbitration lawyers and much larger stakes.

4.5 Backward Induction and Experience

Our experiments involved rather inexperienced subjects (undergrad students compared to professionals). One conjecture is that more experienced players would better understand the incentives at stake, and better perform backward induction. Notice that the performance of our sequential procedures will not decrease and will often increase if players improve at performing backward induction. In VR, on the other hand, a better understanding of the strategic features of the game will not help to resolve the key problem of miscoordination.[15] Our experimental design allows us to test the hypothesis since any

[15]If played multiple times with a same preference profile, then repetition may result in participants coordinating on a specific Nash equilibrium, thereby diminishing the risk of miscoordination. Yet, it is virtually impossible that the same parties would meet on multiple cases with the same set arbitrators and the same preferences over those arbitrators. Instead experience is gained by playing a same procedure with different opponents with different

preference profile is played at an earlier stage by some subjects (rounds 1-10 for $Pf1$ and $Pf4$, and rounds 10-20 for $Pf2$ and $Pf3$) and at a later stage by other subjects (rounds 20-30 for $Pf2$ and $Pf3$, and rounds 30-40 for $Pf1$ and $Pf4$).

Result 5 *There is substantial evidence that participants improve at backward induction with experience, with more observed play paths complying with backward induction when facing a preference profile at a later stage of the experiment.*

SUPPORT: The Supplementary Appendix contains the p-values for testing whether the percentage of play paths that are consistent with backward induction for a given preference profile is statistically different when playing it in an earlier vs. a later rounds. Whenever the difference is statistically relevant (at 5%, and often for much lower thresholds), it goes in the direction that backward induction is played more often in later rounds. Significant differences are obtained for AS with $Pf1$ and $Pf2$, for VAOV with $Pf1$, $Pf3$ and $Pf4$, and for SL with $Pf1$, $Pf2$, and $Pf4$. It is interesting to note that the percentage of social preferences for AS with $Pf2$ and for SL with $Pf3$ (see previous subsection) is virtually identical in earlier vs. later rounds. □

5 Concluding remarks

This paper takes an implementation-theoretic approach to the problem of selecting a public good, namely an arbitrator, to two parties with symmetric information. We first establish that in order to have a mechanism with "socially desirable" properties, one must consider sequential mechanisms and alternative SCRs to the one induced by truthtelling in the commonly used VR procedure. Therefore, we conducted a series of laboratory experiments where we tested four alternative procedures, the VR and three sequential mechanisms that have normatively appealing properties. The experimental analysis yields

preferences over different sets of arbitrators. So skills may improve, but not the level of coordination.

two key results. First, a large fraction of players followed strategic behavior, suggesting that the VR procedure may suffer from the deficiencies outlined in the theoretical section. Second, we find that two *simple* sequential procedures - *which are yet to be used in practice* - the VAOV and SL schemes, are superior to the VR in terms of two criteria: Pareto efficiency and "the minimal satisfaction test".

While our results are presented in the context of arbitrator selection, they potentially may be extended to other situations in which a collective of individuals with symmetric information need to agree on a public good (i.e., an outcome that affects the payoffs of the participants). Examples may include hiring decisions, choosing a set of employees to promote, selecting jury members and deciding on the composition of some committee. Our paper suggests that it may be valuable to study these situations from an implementation-theoretic approach: start by identifying "reasonable" SCRs for the problem at hand; ask whether prevalent procedures implement in theory any of these SCRs; study whether participants in such mechanisms tend to behave according to theory; explore alternative mechanisms that "perform well" both theoretically and behaviorally.

Appendix

Proof of Preliminary Observations. We provide an argument for $n = 5$, the case studied in the experimental section, but it easily generalizes to any n.

Proof of a. Let $A = \{a, b, c, d, e\}$ and (u_1, u_2) generating the following rankings: $a \succ_1 b \succ_1 c \succ_1 d \succ_1 e$ and $b \succ_2 a \succ_2 c \succ_2 d \succ_2 e$ ($Pf2$ in our experiment). Note that reporting truthfully (i.e., vetoing d and e and giving a score of 2 to the top ranked element, a score of 1 to the second-best element and a score of 0 to the remaining element) is not a Nash equilibrium of the veto-rank procedure. If players followed this naïve strategy, they would end up in a tie, where either a or b is randomly chosen. If, on the other hand, player 1 would veto b instead of d, then a would be chosen uniquely, which he prefers.

Consider a preference profile for which a is a Nash equilibrium outcome. Since each player can veto a, this outcome cannot be ranked below the third-best outcome by any player. Consider a pair of strategies (s_1, s_2) with the following properties. Each player i vetoes the elements that player j ranks above a. If the number of elements that j ranks above a is fewer than two, then the remaining elements that i vetoes include elements not vetoed by j (but do not include a). Let B be the set of elements that j vetoes and i does not. If there exists $b \in B$ with $b \succ_i a$, then i reports that a is ranked above b. Otherwise, i reports that every element in B is ranked above a. To see why (s_1, s_2) is a Nash equilibrium note first that no player has a deviation that can lead to an outcome better (for him) than a. Note also that each player's distortion of his true preferences has no effect on the outcome, since this distortion is done with respect to vetoed elements. It follows, that no player can gain by deviating.

Proof of b. Consider first the pair of preference $(\succ_1, \succ_2)$ from the proof of a). Observe that option c, which is Pareto dominated by a and b, is an equilibrium outcome for the previous pair of preferences. On the other hand, one might argue that this Nash equilibrium is not likely to emerge since it involves dominated strategies. Consider then the Bernoulli functions (v_1, v_2) generating the

rankings $a \succ'_1 b \succ'_1 c \succ'_1 d \succ'_1 e$ and $e \succ'_2 c \succ'_2 a \succ'_2 b \succ'_2 d$ ($Pf4$ in our experiment). Then $f_{VR}(\succ'_1, \succ'_2) = \{a\}$. However, there exists a (undominated) Nash equilibrium in which player 2 chooses $V_2 = \{a, b\}$ and s_2 such that $s_2(e) = 2$, $s_2(c) = 1$ and $s_3(d) = 0$, while player 1 chooses $V_1 = \{d, e\}$ and s_1 such that $s_1(a) = 2$, $s_1(b) = 1$ and $s_1(c) = 0$. The outcome of this equilibrium is c, which does not belong to $f_{VR}(\succ'_1, \succ'_2) = \{a\}$.

Proof of c. The preference profile $(\succ_1, \succ_2)$ described in the proof of a) induces a pair of equilibria, s and s', with the following properties. In s player 2 is truthful, while player 1 chooses $V_1 = \{b, e\}$ and $r_1(a) = 2$, $r_1(c) = 1$ and $r_1(d) = 0$. In s' player 1 is truthful, while player 2 chooses $V'_1 = \{a, e\}$ and $r_1(b) = 2$, $r_1(c) = 1$ and $r_1(d) = 0$. It follows that (s_1, s'_2) induces the Pareto dominated outcome c. ■

Proof of Proposition 1 Part a) follows as a Corollary of Hurwicz and Schmeidler (1978). Indeed, they proved that any SCR that is Pareto efficient and partially implementable must be dictatorial. Any such SCR will thus fail the MST.

We now pay attention to RSCFs. The proof is made for the case where A contains five elements - $A = \{a, b, c, d, e\}$ - but can easily be extended to any number of elements. Consider (u_1, u_2) such that $u_1(a) > u_1(b) > u_1(c) > u_1(d) > u_1(e)$, and u_2 is completely opposite. If ψ passes the MST, then $\psi(u_1, u_2)$ yields c with certainty. Maskin Monotonicity implies that $\psi(u'_1, u'_2)$ also yields c with certainty, where $u'_1(c) > u'_1(e) > u'_1(a) > u'_1(b) > u'_1(d)$ and $u'_2(e) > u'_2(c) > u'_2(a) > u'_2(b) > u'_2(d)$. Consider (u''_1, u''_2) such that $u''_1(c) > u''_1(a) > u''_1(e) > u''_1(b) > u''_1(d)$, and u''_2 is completely opposite. If ψ passes the minimal satisfaction test, then $\psi(u''_1, u''_2)$ yields e with certainty. Maskin monotonicity then implies that $\psi(u'_1, u'_2)$ also yields e with certainty, a contradiction. This establishes b).

Statement c) then follows from a) and b), given that ψ_{VR} and f_{VR} are Pareto efficient and satisfy the MST. ■

Proof of Proposition 2 Suppose, to the contrary of what we want to prove, that there exists an extensive-form mechanism that leads to backward induc-

tion outcomes that systematically fall within f_{VR}. Let a, b, c, d be four elements of A. Consider the following pair $(\succ_1, \succ_2)$ of orderings: $a \succ_1 b \succ_1 c \succ_1 d \succ_1 x$ and $d \succ_2 c \succ_2 a \succ_2 b \succ_2 x$, for each $x \in A \setminus \{a, b, c, d\}$. The backward induction outcome computed for this pair of preferences must be a, since this is the only element in $f_{VR}(\succ_1, \succ_2)$. Let now $\succ_2'$ be the same preference ordering as $\succ_2$, except that the relative ranking of c and d is reversed. We now prove that a must be the backward induction outcome of the mechanism when computed for $(\succ_1, \succ_2')$. For each decision node ν, let $\mathcal{O}(\nu, \succ_1, \succ_2)$ be the set of arbitrators that the party in charge at ν can generate by choosing various actions, while assuming that the rest of the extensive-form will be played by backward induction for $(\succ_1, \succ_2)$. A similar construction defines $\mathcal{O}(\nu, \succ_1, \succ_2')$. We prove by backward induction that $\mathcal{O}(\nu_1, \succ_1, \succ_2) \cap \{a, b\} \neq \emptyset$ if and only if $\mathcal{O}(\nu_1, \succ_1, \succ_2') \cap \{a, b\} \neq \emptyset$, for each decision node ν_1 at which the first party makes a choice, and $\mathcal{O}(\nu_2, \succ_1, \succ_2) \cap \{c, d\} \neq \emptyset$ if and only if $\mathcal{O}(\nu_2, \succ_1, \succ_2') \cap \{c, d\} \neq \emptyset$, for each decision node ν_2 at which the second party makes a choice. This is trivially true if these are the last decision nodes. Consider then a decision node ν_1 where the first party makes a decision, and suppose that the property holds true at every subsequent node. We may assume without loss of generality that all the nodes that come right after a decision from the first party are nodes where the second party makes a decision. The second party's optimal action at those nodes leads to an element of $\{c, d\}$ if there is an action that leads to one of these two outcomes when the rest of the subgame is played by backward induction for either $(\succ_1, \succ_2)$ or $(\succ_1, \succ_2')$ (which ever pair of preferences is used to express this condition is irrelevant, thanks to the induction hypothesis). The optimal action at the other nodes does not change when moving from $\succ_2$ to $\succ_2'$ and vice versa, since $\{c, d\}$ is inaccessible for $(\succ_1, \succ_2)$ if and only if it is inaccessible for $(\succ_1, \succ_2')$ (by the induction hypothesis). Hence $\mathcal{O}(\nu_1, \succ_1, \succ_2) \cap \{a, b\} \neq \emptyset$ if and only if $\mathcal{O}(\nu_1, \succ_1, \succ_2') \cap \{a, b\} \neq \emptyset$, as desired. A similar argument applies for a decision node ν_2 at which the second party makes a decision. Take now a node that is reached when the equilibrium strategies for $(\succ_1, \succ_2)$ are followed. The first party has an action that leads to a if the equilibrium path is followed

thereafter. This is the best possible option for him, so he has no incentive to take any alternative action if the equilibrium path is followed thereafter, and this is independent of what happens in the subgames that would be reached if he were to choose a different action. The second party also has an action that leads to a if the equilibrium path is followed thereafter. If there was an action that would lead to either c or d when the rest of the game is played thereafter according to the backward induction strategies for $(\succ_1, \succ_2')$, then there would be one that would also lead to c or d when backward induction is applied to $(\succ_1, \succ_2)$ instead, thanks to the property we just proved. No such action exist for the second party along the equilibrium path for $(\succ_1, \succ_2)$, and hence it must be that the second party's action remains optimal when his preference is $\succ_2'$ instead of $\succ_2$, while taking into account that the rest of the game will be played by backward induction according to $(\succ_1, \succ_2')$. Arbitrator a is thus the backward induction outcome of the extensive-form mechanism for $(\succ_1, \succ_2')$. Let $\succ_1'$ be the ordering that coincides with $\succ_1$, except that the ordering of a and b are reversed. The backward induction equilibrium computed for this pair of preferences $(\succ_1', \succ_2')$ must be c, since this is the only element in $f_{VR}(\succ_1', \succ_2')$. A similar reasoning to the one used to move from $(\succ_1, \succ_2)$ to $(\succ_1, \succ_2')$ will imply that c must also be a backward induction outcome of the extensive-form mechanism for $(\succ_1, \succ_2')$. This leads to a contradiction since extensive-form games have a unique backward induction outcome given strict preferences. ■

Proof of Proposition 4 It is easy to check that the two-stage shortlisting mechanism proposed implements o^*, and that o^* is Pareto efficient and passes the MST. Hence we will limit ourselves to prove that o^* is the only SCR with those properties. Let $\succ \in \mathcal{P} \times \mathcal{P}$. We now define a new ordering $\succ_1'$ for the first mover. First the elements ranked above $o^*(\succ)$ according to $\succ_1$ keep the same rank[16] in $\succ_1'$. Notice that the rank of all these elements must be strictly larger than $\frac{n+1}{2}$ in $\succ_2$, by definition of o^*. Then place the other elements ranked strictly larger than $\frac{n+1}{2}$ in $\succ_2$ (if any) in some specific order (let's say

[16]The rank of a top ranked element is 1. The rank of the second element according to the ordering is 2, and so on so forth. The rank is thus equal to n minus the score.

alphabetically) in the next available spots in $\succ_1'$ (that is, after those elements above $o^*(\theta)$ according to $\succ_1$). The next available spot in $\succ_1'$ must be the $\frac{n+1}{2}$-rank. Place $o^*(\theta)$ there, and then rank the remaining elements in some specific order (let's say alphabetically again). Let $\bar{o}$ be a single-valued SCR that can be implemented via a two-stage mechanism, is Pareto efficient and passes the MST. The MST applied to both players implies that $\bar{o}(\succ_1', \succ_2) = o^*(\succ)$. Notice that the lower contour set of $o^*(\succ)$ expands when moving from $\succ_1'$ to $\succ_1$. Hence the backward induction outcome of the two-stage mechanism in $(\succ_1, \succ_2)$ must be the same as the one in $(\succ_1', \succ_2)$ (the second party's optimal strategy remains unchanged since his preference remains fixed), or $\bar{o}(\succ) = \bar{o}(\succ_1', \succ_2)$. We get $\bar{o}(\succ) = o^*(\succ)$ by transitivity, as desired. ■

References

Abdulkadiroglu, A., P. A. Pathak, and A. E. Roth, 2005a. The New York City High School. *American Economic Review (Papers and Proceedings)* **95**, 364-367.

Abdulkadiroglu, A., P. A. Pathak, A E. Roth, and T. Sönmez, 2005b. The Boston Public School. *American Economic Review (Papers and Proceedings)* **95**, 368-371.

Anbarci, N., 2006. "Finite Alternating-Move Arbitration Schemes and the Equal Area Solution." *Theory and Decision* **61**, 21-50.

Bierbrauer, F. and N. Netzer, 2012. "Mechanism Design and Intentions." University of Zurich Working Paper.

Binmore, K., J. McCarthy, G. Ponti, L. Samuelson, and A. Shaked, 2002. A Backward Induction Experiment. *Journal of Economic Theory* **104**, Pages 48-88.

Bloom, D. E., and C. L. Cavanagh, 1986a. An Analysis of the Selection of Arbitrators. *American Economic Review* **76**, 408-422.

Bloom, D. E., and C. L. Cavanagh, 1986b. An Analysis of Alternative Mechanisms for Selecting Arbitrators. Harvard Institute of Economic Research Discussion Paper 1224.

Bowles, S. and S-H Hwang, 2011. "The Sophisticated Planner's Dilemma: Mechanism Design with Endogenous Preferences." Sante Fe Institute Working Paper.

Brams, S. J. and D. M. Kilgour, 2001. "Fallback Bargaining." *Group Decision and Negotiation* **10**, 287-316.

Charness, G. and M. Rabin, 2002. "Understanding Social Preferences with Simple Tests." *Quarterly Journal of Economics* **117**, 817-869.

Chen, Y., 2008. Incentive-Compatible Mechanisms for Pure Public Goods: A Survey of Experimental Research. In C. Plott and V. L. Smith, eds., The Handbook of Experimental Economics Results, Amsterdam: Elsevier Press, 625-643.

Crawford, V. P., M. Costa-Gomes, and N. Iriberri, 2010. "Strategic Thinking". Mimeo, UCSD.

Degan, A., and A. Merlo, 2009. Do voters vote ideologically? *Journal of Economic Theory* **144**, 1868-1894.

Forsythe, R., R. B. Myerson, T. A. Rietz, and R. J. Weber, 1993. An Experiment on Coordination in Multi-Candidate Elections: The Importance of Polls and Election Histories. *Social Choice and Welfare* **10**, 223-247.

Forsythe, R., R. B. Myerson, T. A. Rietz, and R. J. Weber, 1996. An Experimental Study of Voting Rules and Polls in Three-Candidate Elections. *International Journal of Game Theory* **25**, 355-383.

Hurwicz, L, and D. Schmeidler, 1978. Construction of Outcome Functions Guaranteeing Existence and Pareto Optimality of Nash Equilibria. *Econometrica* **46**, 1447-1474.

Hurwicz, L., and M. R. Sertel, 1997. Designing Mechanisms, in particular for Electoral Systems: The Majoritarian Compromise. Mimeo.

Kagel, J. H., 1995. Auctions: A Survey of Experimental Research. In J. H. Kagel and A. E. Roth, eds., Handbook of Experimental Economics. Princeton, NJ: Princeton University Press, 501-585.

Kagel, J. H., and D. Levin, 2011. Auctions: A Survey of Experimental Research, 1995-2010. Handbook of Experimental Economics, Vol. 2., *forthcoming.*

Kawai, K., and Y. Watanabe, 2010. Inferring Strategic Voting. Mimeo, Northwestern University.

Kibris, O. and M. R. Sertel, 2007. "Bargaining Over a Finite Set of Alternatives." *Social Choice and Welfare* **28**, 421-437.

Levitt, S. D., J. A. List, and S. E. Sadoff, 2011. "Checkmate: Exploring Backward Induction Among Chess Players." *American Economic Review* **101**, 975–990

Roth, A. E., 1984. The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory. *Journal of Political Economy* **92**, 991-1016.

Roth, A. E., 2002. The Economist as Engineer: Game Theory, Experimental Economics and Computation as Tools of Design Economics. *Econometrica* **70**, 1341-1378.

Roth, A. E., 2007. The Art of Designing Markets. *Harvard Business Review*, October Issue, 118-126.

Sprumont, Y., 1993. "Intermediate Preferences and Rawlsian Arbitration Rules." *Social Choice and Welfare* **10**, 1-15.